CONTEXT-AWARE Q&A ON SPACE DATA

This project enables automated understanding and answering of space-related questions using a context-aware model trained on real datasets. It mimics how humans extract relevant information from passages, helping improve educational tools, research assistance, and public access to complex space knowledge. By learning question-answer patterns, the system enhances interaction with large text datasets and supports scalable knowledge extraction.

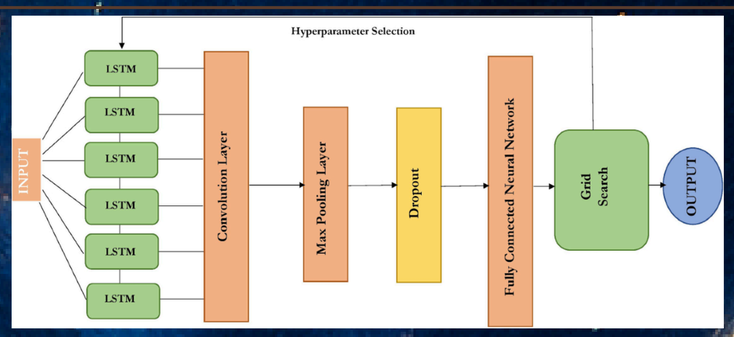
Definition

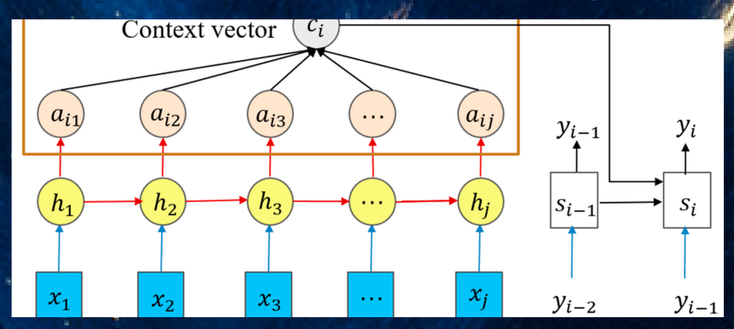
An encoder-decoder model transforms input text into meaningful output text. The encoder understands the input and creates a context; the decoder uses this context to generate a relevant response.

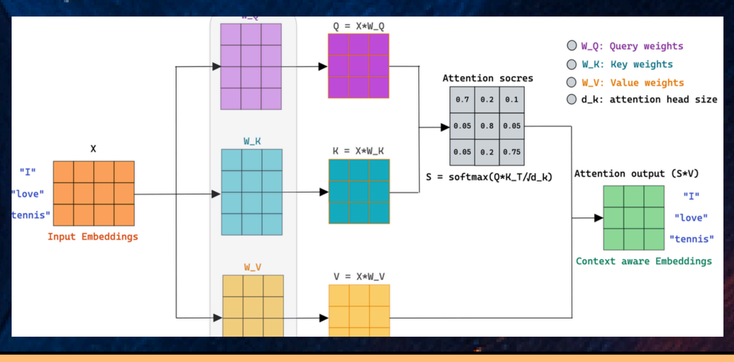
1. Introduction

Question Answering (QA) systems are a fundamental building block for achieving human-like interaction with text-based information. In this project, we concentrate on extractive QA, where a distinct chunk needs to be extracted from a provided context to supply an answer to a question. We employ a tailored SQuAD-type dataset centered on the domain of space science and technology, where each example is a context passage, question, and corresponding answer. Our effort is to evaluate the performance of three distinct architectures on this task:

- Baseline LSTM without attention
- LSTM with Bahdanau attention manual
- Transformer with self-attention (no recurrence)







2. Methodology

Our pipeline involved structured preprocessing, custom model architectures, and GPU-based training with standardized evaluation.

- Custom SQuAD-style JSON with 1000+ context-question-answer triplets from the space domain.
- Tokenization using Keras' Tokenizer
- Padding sequences to fixed lengths for context, question, and answer
- Decoder inputs created with <start> token; outputs shifted with <end>

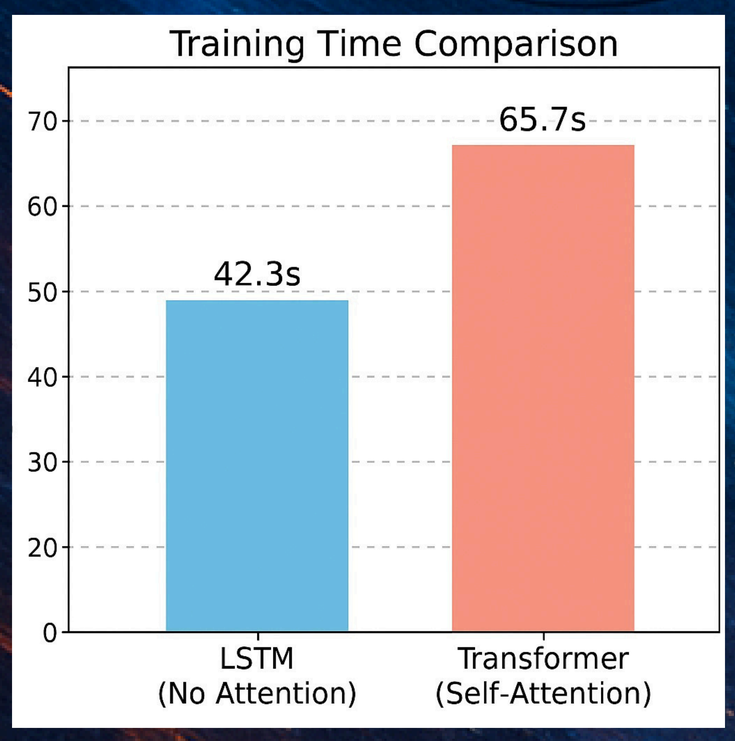
◆ Model Architectures

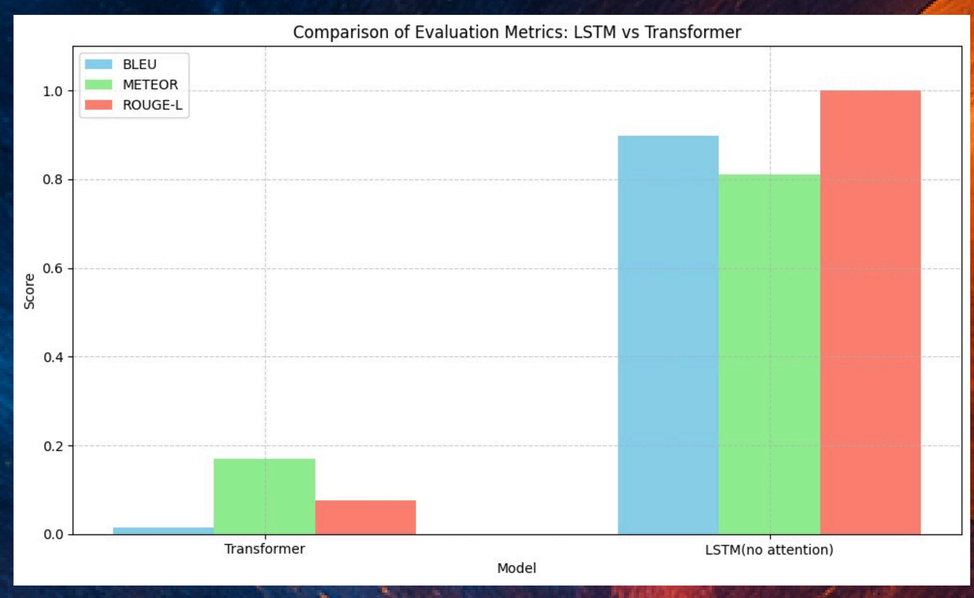
- LSTM Baseline: Simple encoder-decoder with shared embeddings.
- LSTM + Bahdanau Attention: Adds soft attention over encoder outputs.
- Transformer: Encoder-decoder using multi-head self-attention and positional encodings.

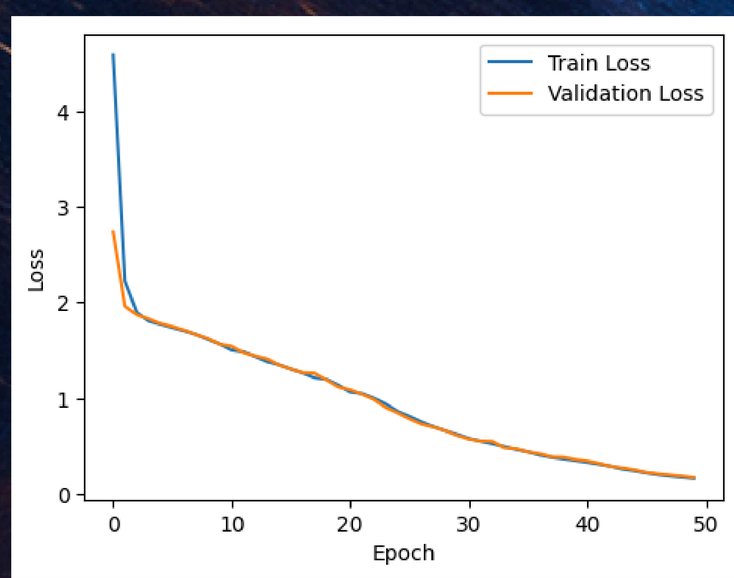
◆ Training Setup

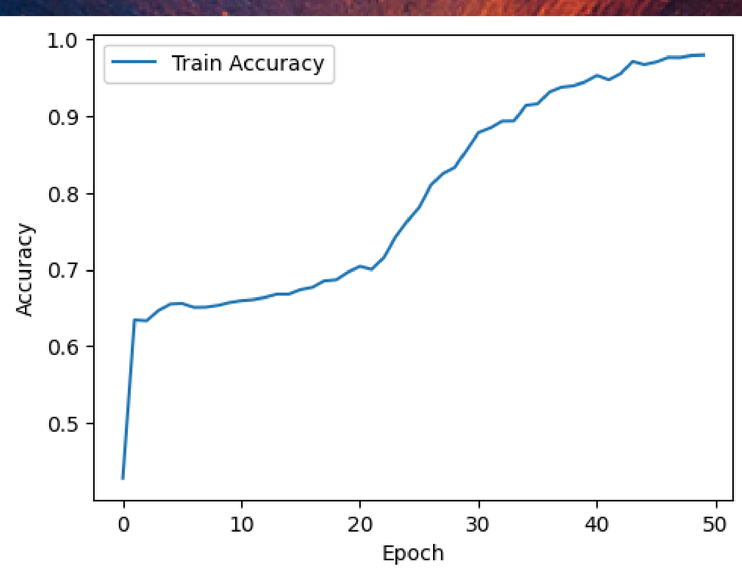
- Loss: sparse_categorical_crossentropy
- Optimizer: Adam

3. Results









4. Conclusion

This work demonstrates the evolution from the classic recurrent models to attention-based models for extractive QA. Although the LSTM baseline is reasonable enough, Bahdanau attention enhances contextual alignment, and Transformers dominate both by considering full input dependencies. These results confirm the effectiveness of attention in contemporary QA systems, particularly where accuracy and explainability are important in domain-specific applications such as space science.

Presented By-

Pranav Khatavkar - 202201090028 Mech

Aditya Sawleshwarkar - 202201090045 Mech

Omkar Karlekar - 202201090088 Mech

Guided By-

Diptee Chikmurge